Accepted Manuscript

British Journal of General Practice

Identification of symptoms associated with the diagnosis of pancreatic exocrine and neuroendocrine neoplasms: a nested case-control study of the UK population

Liao, Weiqi; Clift, Ashley; Patone, Martina; Coupland, Carol; González-Izquierdo, Arturo; Pereira, Stephen; Hippisley-Cox, Julia

DOI: https://doi.org/10.3399/BJGP.2021.0153

To access the most recent version of this article, please click the DOI URL in the line above.

Received 04 March 2021
Revised 28 June 2021
Accepted 02 July 2021

Identification of symptoms associated with the diagnosis of pancreatic exocrine and neuroendocrine neoplasms: a nested case-control study of the UK population

Weiqi Liao MMed, MRes, Data Scientist [1], Ashley K Clift MA, MBBS, clinical research fellow [1, 2], Martina Patone PhD, Data Scientist [1], Carol Coupland PhD, professor of medical statistics in primary care [3], Arturo González-Izquierdo PhD, Senior Research Associate [4], Stephen P Pereira MD, Professor of Hepatology and Gastroenterology [5], Julia Hippisley-Cox MD, FRCGP, FRCP, Professor of Clinical Epidemiology and General Practice [1]

1. Nuffield Department of Primary Care Health Sciences, University of Oxford
2. Cancer Research UK Oxford Centre, University of Oxford
3. Division of Primary Care, School of Medicine, University of Nottingham
4. UCL Institute of Health Informatics, Health Data Research UK, London
5. UCL Institute for Liver and Digestive Health, University College London

Corresponding author:

Professor Julia Hippisley-Cox

Nuffield Department of Primary Care Health Sciences, University of Oxford

Radcliffe Observatory Quarter, Woodstock Road, OX2 6GG, Oxford

julia.hippisley-cox@phc.ox.ac.uk

Running head: Symptoms associated with PDAC and PNEN

Abstract

Background: Pancreatic cancer has the worst survival rate among all cancers. Almost 70% of patients were diagnosed at Stage IV.

Aim: This study aimed to investigate the symptoms associated with the diagnoses of pancreatic ductal adenocarcinoma (PDAC) and neuroendocrine neoplasms (PNEN), comparatively characterise the symptomatology between the two tumour types to inform earlier diagnosis.

Design and Setting: A nested case-control study was conducted using data from the QResearch database. Patients aged ≥25 years and diagnosed with PDAC or PNEN during 2000-2019 were the cases. Up to 10 controls from the same general practice were matched with each case by age, sex, and calendar year using incidence density sampling.

Methods: Conditional logistic regression was used to investigate the association between the forty-two shortlisted symptoms and the diagnoses of PDAC/PNEN in different timeframes relative to the index date, adjusting for patients' sociodemographic characteristics, lifestyle, and relevant comorbidities.

Results: There were 23,640 patients diagnosed with PDAC and 596 with PNEN. Twenty-three symptoms were significantly associated with PDAC, and nine symptoms with PNEN. Jaundice and gastrointestinal bleeding were the two alarm symptoms for both tumours. Thirst and dark urine were the two new identified symptoms for PDAC. The risk of unintentional weight loss may be longer than two years before the diagnosis of PNEN.

Conclusion: PDAC and PNEN have overlapping symptom profiles. The QCancer (Pancreas) risk prediction model could be updated by including the newly identified symptoms and comorbidities, which could help GP identify high-risk patients for timely investigation in primary care.

Keywords

Pancreatic Neoplasms [MeSH], early diagnosis [MeSH], primary health care [MeSH], pancreatic ductal adenocarcinoma (PDAC), pancreatic neuroendocrine neoplasms (PNEN), symptom

How this fits in

This is the largest population-based study of its kind, systematically examining the symptomology of pancreatic ductal adenocarcinoma (PDAC) and pancreatic neuroendocrine neoplasms (PNEN), and quantifying the association of 42 potential symptoms in different time windows relative to the date of diagnosis. This study confirmed several symptoms as risk factors for PDAC reported in previous UK studies with much smaller sample sizes. A deeper understanding of symptoms associated with PNEN is gained beyond archetypal but rare symptoms of hormone hypersecretion.

Considering most symptoms associated with pancreatic cancer are non-specific and not qualified for urgent referral for investigation (two-week wait) in the current NICE guideline, general practitioners (GP) should be vigilant of patients presenting with several concurrent non-specific symptoms and make proper safety-netting strategies. GP should also increase the awareness of the risk of pancreatic cancer among people with comorbidities, and be careful not to attribute potential symptoms of pancreatic cancer to patients with existing health conditions.

Introduction

Pancreatic cancer is the tenth most common cancer in incidence but represents the fifth most common cause of cancer death in the UK. Pancreatic cancer is very aggressive and has the worst survival rate among all types of cancer (1). Tumours arising from the pancreas can be classified as exocrine (approximately 95%) or neuroendocrine (≤5%) neoplasms (2), known as pancreatic ductal adenocarcinoma (PDAC) and pancreatic neuroendocrine neoplasms (PNEN) (3), respectively. Although there are differences in tumour pathobiology and treatment strategies between PDAC and PNEN, the two tumour types share a proclivity to metastasis (4-6). More favourable outcomes were observed in earlier stages at diagnosis (7).

In the absence of a screening programme for pancreatic cancer in the UK, symptomatic presentation in general practice remains a key avenue for earlier diagnosis. However, besides jaundice, current literature reporting symptoms associated with pancreatic cancer are vague and non-specific (8, 9). General practitioners (GPs) face the challenges of differentiating a potential malignancy from other benign diseases when patients present with non-specific symptoms, which could be easily 'missed' or delayed diagnosis of the tumour. In addition, as a rarer type of cancer, there is still no large scale study characterising the symptomatology of PNEN, nor have studies comprehensively compared the symptomatology of PDAC with PNEN. Therefore, a better understanding of the symptomatology and the timings at which patients present with symptoms could help GPs better manage patients and make clinical decisions. Furthermore, such symptoms could be used in social campaigns to increase public awareness of pancreatic cancer.

To address this research gap, we conducted this study with three aims: 1) to explore the symptoms that patients presented in primary care in different time windows that may portend the diagnosis of PDAC/PNEN; 2) to comparatively characterise the symptomatology of PDAC and PNEN, and 3) to inform the update of the QCancer (Pancreas) prediction model (10, 11). The QCancer (Pancreas) score quantifies the risk of an incident diagnosis of pancreatic cancer in the next two years, based on an individual patient's characteristics. Such a risk score may help GP make different decisions: a two-week wait referral for patients with high-risk, watchful wait and safety-netting for patients with low risk.

Methods

Study design and setting

This is a nested case-control study using the QResearch database (Version 44), an extensive validated anonymised primary care database comprising records of over 35 million patients registered in approximately 1500 GP surgeries spread throughout the UK using the EMIS System since 1989. Population in the QResearch database is representative of the UK population. Patients' records have been linked with cancer registries, ONS mortality records, and hospital episode statistics. The cancer registration data include information on the date of diagnosis, type and location of the tumour, morphology, grade and stage, and treatment.

Study population

An open cohort of patients aged ≥25 years and registered in the QResearch database between 1 January 2000 and 31 December 2019 was the eligible study population. Patients with an existing diagnosis of any type of pancreatic cancer before the entry date were excluded. The entry date to the cohort was the latest of the patient's 25th birthday, the date of patients registered with the practice plus one year, the date on which the practice computer system was installed plus one year, or the beginning of the study period. The right censor date was the earliest date of the following: the date of pancreatic cancer diagnosis, the date of death, the date leaving the practice, or the study end date. Person years were calculated between the study entry date and the right censor date.

Identification of cases and controls

Cases were patients in the study cohort with an incident diagnosis of PDAC/PNEN, recorded in one or more of the four linked sources – GP records, HES, cancer registry, or ONS. The index date for cases was the earliest date the diagnosis was recorded in any four data sources. Cases were matched with up to 10 controls in the same practice, age, sex, and calendar year using incidence density sampling (12). Each control was allocated an index date, which was the date of diagnosis of their matched case.

Candidate symptoms and potential risk factors

A broad list of symptoms potentially associated with PDAC and PNEN is summarised in Box 1. These symptoms were identified through literature review (2, 10, 13, 14), information from the leading charities such as Cancer Research UK (15) and Pancreatic Cancer UK (16), NICE guidelines – NG12 (17) and NG85 (18), and patient representatives. All occurrences of symptoms in primary care records were extracted, but the analysis was focused on the most recent five years before the index date. In addition, the following variables were of research interest and adjusted in the models, including patients' sociodemographic characteristics (ethnicity, socioeconomic deprivation using Townsend quintile), lifestyle factors (smoking and drinking statuses, and body mass index (BMI), using most recent available values before the index date), and relevant comorbidities that could cause the symptoms, or as potential risk factors for pancreatic cancer.

Statistical analysis

Descriptive statistics were used to summarise the sociodemographic and clinical characteristics of patients diagnosed with PDAC and PNEN, and the matched control group. The key clinical characteristics of PDAC and PNEN cases were compared.

Exploratory analyses were conducted to investigate the association between the most recent recorded symptom relative to the index date in seven different periods and patient groups (case/control) using univariable conditional logistic regression. These seven timeframes were <1 month, 1-3 months, 4-6 months, 7-12 months, 1-2 years, 2-3 years, and 3-5 years before the index date. The purpose of setting different timeframes for the same symptoms was to compare how the odds ratio (OR) would change, and the implication of timeframes for earlier diagnosis of PDAC and PNEN based on symptomatic presentation. Based on the exploratory results and the clinical relevance of timeframes for earlier diagnosis of pancreatic cancer, the seven timeframes were narrowed down to the following four in two sets of analysis:

1) Within 3 months for alarm symptoms, or 1 year for other symptoms (denoted as 3M/1Y);
2) Within 6 months for alarm symptoms, or 2 years for other symptoms (6M/2Y)

Alarm symptoms included jaundice, dysphagia, and gastrointestinal (GI) bleeding. These three symptoms were analysed within shorter timeframes (within 3/6 months before the index date), as they are widely accepted as 'red flag' symptoms and should be promptly investigated in primary care, or referred to secondary care. For the other symptoms, the timeframes were longer (within 1 or 2 years before the index date), as they are non-specific, probably caused by other benign conditions, and not easily ascribed to an underlying tumour. A categorical variable was used to denote whether the patients presented to their GPs for each symptom based on the most recent date of presentation, for example, no record of presenting with jaundice (reference category), presenting with jaundice within 3 months, or more than 3 months before the index date. Symptoms in other timeframes (6 months, 1 or 2 years) were operationalised in the same way.

Due to the large difference in sample sizes, PDAC and PNEN were analysed separately. Five variables contained missing data, including ethnicity, Townsend quintile, BMI, smoking and drinking statuses. Multiple imputation with chained equations was used to impute missing values for these variables under the missing at random assumption. Ten imputations were conducted. Multivariable conditional logistic regression models were used to identify symptoms significantly associated with the diagnoses of PDAC and PNENs, adjusting for patient characteristics and comorbidities, with Rubin's rules used to pool the parameter estimates across the ten imputed datasets (19). Possible interactions were considered and tested in the model. ORs and the 95% confidence intervals (CI) for each symptom were calculated and visualised in forest plots. Symptoms with an OR >1.2 at a significance level of $P<0.01$ for PDAC or PNEN were considered clinically and statistically relevant. Sensitivity analyses were conducted in patients (both cases and control) with at least three years of electronic health records (EHRs) before the index date. All statistical analyses were conducted in Stata 16.1. The reporting of this study followed the recommendations of the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement (20).

Results

Population characteristics

The open cohort included 15,194,279 patients aged ≥25 years, with a total of 100,290,294 person-years of follow-up. 23,640 PDAC and 596 PNEN cases were identified from the cohort. Case ascertainment from the linked data sources is in Supplementary Table 1. The age-standardised incidence rate of PDAC was 23.21 per 100,000 person-years (95% CI 22.91-23.51) and 0.95 per 100,000 person-years for PNEN (95% CI: 0.89-1.01). Table 1 summarises the demographic and clinical characteristics of the cases, organised by tumour type. The comparison of characteristics between cases and controls (n=230,024) is in Table 2.

Symptoms associated with PDAC and PNEN

Figures 1 presents the significant results of multivariable analyses for PDAC and PNEN for symptoms recorded within 3 months (alarm symptoms) or 1 year (non-specific) before the index date, respectively, adjusting for patient characteristics and comorbidities. The significant results for symptoms in the 6M/2Y timeframe for PDAC and PNEN are in Figure 2 (considering the publication space, we only plotted the significant results, but full results are available in Supplementary Tables 2-5). Most results were congruent in the two models, although symptoms in a shorter timeframe (3M/1Y) generally had higher ORs than those in a longer period (6M/2Y) and had wider confidence intervals. In addition, symptoms within the cut-off periods (e.g. 3/6 months or 1/2 years) were statistically significant. For symptoms longer than the cut-off periods, there were two main patterns: either the symptoms became non-significant, or the direction of OR in the symptoms reversed (from >1 to <1, significantly higher odds in control), which meant the controls were more likely to consult with those (non-specific) symptoms after the cut-off periods. Noticeably, the effect of weight loss may be longer than 2 years before the diagnosis of PNEN. Interaction terms were tested. However, they were not in the final model because including interaction terms increased the number of parameters that need to be estimated. The model was unable to converge, especially for PNEN with a small sample size.

Jaundice had the highest adjusted OR in both PDAC and PNEN. Symptoms associated with PNEN were a subset of symptoms associated with PDAC, but the strength of ORs among the significant symptoms may not be the same in PDAC and PNEN. Nine symptoms were significantly associated with the diagnosis of PNEN in the timeframe of 3M/1Y, including jaundice, GI bleeding, diarrhoea, bowel change, vomiting, indigestion, abdominal mass, abdominal pain, and weight loss. The additional significant symptoms associated with PDAC included constipation, steatorrhea, abdominal distension, nausea, flatulence, heartburn, fever, tiredness, appetite loss, itching, back pain, thirst, and dark urine.

Other risk factors

Compared with white ethnicity, Indian, Bangladeshi, other Asian (not including Chinese) were less likely to develop PDAC (OR<1). Smoking and drinking were the risk factors for PDAC. As to comorbidities, type 2 diabetes mellitus (T2DM), venous thromboembolism, Cushing's syndrome, pancreatic cysts, significantly increased the risks of PDAC and PNEN. Acute pancreatitis, cholangitis, family history of GI cancer, and type 1 diabetes were significant risk factors for PDAC, but not for PNEN.

Sensitivity analysis

About 26% of patients (65,884 out of 254,260, including cases and controls) were excluded from sensitivity analysis. By comparing the results from the main analysis, the conclusion of symptoms in the sensitivity analysis did not change, although there were some changes in OR. Complete results of sensitivity analyses are available in Supplementary Tables 6-9.

Discussion

Summary

24,236 patients diagnosed with pancreatic cancer during 2000-2019 were identified from the QResearch database. Nine symptoms were significantly associated with the diagnosis of PNEN, which is a subset of 23 significant symptoms for PDAC. A shorter timeframe (3 months for alarm symptoms and 1 year for non-

specific symptoms) was considered better than a longer one (6M/2Y) for earlier cancer diagnosis, as cases had higher odds of presenting symptoms in the 3M/1Y timeframe. Jaundice had the highest adjusted OR in both PDAC and PNEN. Thirst and dark urine were the two newly identified symptoms associated with PDAC, not previously reported in other studies. Thirst could be a symptom explained by T2DM, which is associated with pancreatic cancer. Dark urine could be caused by progressing liver dysfunction, or the manifestation of biliary duct obstruction. All the findings in this study are summarised in Table 3.

Strengths and limitations

The QResearch database provides rich data for this study, which is by far the largest study of its kind. The representative patient population makes the study findings more able to generalise to a broader UK population. The use of EHRs avoids selection, recall, and responder biases from survey, also benefits from the accuracy of coding and data completeness in the UK general practice. We explored the effects of symptoms in seven timeframes first, and then narrowed down to 3M/1Y and 6M/2Y before the index date. We did not mix symptoms recorded longer than the timeframes (>3M/1Y, >6M/2Y) with no symptom recorded together, which provided new information about the symptoms beyond the cut-off periods in cases and controls. Although symptoms were the focus of this study, we took account of, included, and adjusted for the background risk factors and relevant comorbidities in the model, which is another strength. We conducted the study as transparently and thoroughly as possible. The research protocol has been published on the QResearch website. The reporting of this article complies with the STROBE statement.

Information bias in EHRs is the first limitation. We could not evaluate how accurate the information was recorded across practices. The recording habit may have a considerable difference among GPs. We mitigated the heterogeneity of recording habits by using all possible Read codes for each variable. Due to the small sample size in PNEN cases, it is possible that we could not capture the full burden of symptoms in PNEN. Therefore, we could not discern whether a lower number of significant symptoms for PNEN is a lack of statistical power, or PNEN has truly less prominent symptomatology, or both. We planned to explore whether there was any symptom associated with early/late stages at diagnosis in PDAC and PNEN. Unfortunately, the large amount of missing data in cancer staging (Supplementary Table 10) did not allow us to conduct such an analysis.

Comparison with existing literature

The current study has some improvements from our previous QCancer (Pancreas) prediction model (10), including a longer study period, a larger sample size of incident cases, more symptoms examined, and the exploration of PNEN, which resulted in additional thirteen significant symptoms identified from this study. Some UK studies examined the symptomatology of PDAC in primary care settings, with a similar study design (matched case-control study) and statistical method (conditional logistic regression), using the GPRD (21) and the Health Improvement Network (THIN) (22) databases. The findings in this study are generally consistent with the two publications. No studies have systematically and robustly evaluated the symptomatology of PNEN in primary care. Patients with PNEN (n=64) reported their symptoms in a voluntary, internet-based survey. Given the small sample size and potential recall bias in that study (23), we believe our population-based approach offers a more robust and generalisable insight of the PNEN symptomatology. Older age, smoking, excess alcohol intake, chronic pancreatitis, and T2DM are common risk factors for pancreatic cancer (24-28). The findings are the same here.

Implications for practice and research

Most symptoms identified in this study do not qualify for a rapid referral in the current NICE guideline for suspected (pancreatic) cancer pathway referral [NG12](17). GPs should be vigilant to patients presenting with alarm symptoms and non-specific but concerning symptoms, especially when patients have existing comorbidities. Public and patient engagement events could raise public awareness of the symptoms of pancreatic cancer, which may help patients see their GPs more promptly when noticing bodily changes.

Based on the study findings, we can update the QCancer (Pancreas) prediction model (10) and develop a new model for PNEN. It is also possible to quantify the risk of patients presenting with several concurrent non-specific symptoms, and the predictive values of such symptom combinations. It would be interesting to further understand how GPs managed and investigated patients presented with different symptom combinations, and the association with the route to diagnosis and cancer stage. In addition, there is an ongoing project in our team, investigating diabetes as a risk pathway towards pancreatic cancer

(https://www.qresearch.org/research/approved-research-programs-and-projects/diabetes-as-a-risk-pathway-towards-early-diagnosis-and-prognostication-of-pancreatic-cancer/).

Conclusion

Early diagnosis of pancreatic cancer from primary care is still challenging, due to the non-specific symptoms. This study identified 23 symptoms associated with the diagnosis of PDAC and nine symptoms for PNEN. Risk prediction models incorporating comprehensive symptomatology would help identify patients with a high risk of developing pancreatic tumours from primary care. Patients could benefit from an earlier cancer diagnosis and better survival outcomes, which can also save costs for the NHS.

Acknowledgements

This project involves data from patient-level information collected by the NHS, as part of the care and support of patients. We acknowledge the practices that contribute to EMIS (Egton Medical Information Systems) Health and the QResearch database, and the Universities of Nottingham and Oxford for the expertise in establishing, developing, and supporting the QResearch database. The Hospital Episode Statistics data used in this study are re-used with permission from NHS Digital, who retain copyright on the data. The authors thank the Office for National Statistics for providing the mortality data.

Funding

This study was funded by the Pancreatic Cancer UK Early Diagnosis Award 2018, project title: The Accelerated Diagnosis of neuroEndocrine and Pancreatic TumourS (ADEPTS). QResearch received funding from the National Institute for Health Research Biomedical Research Centre, Oxford, grants from John Fell Oxford University Press Research Fund, grants from Cancer Research UK (Grant number C5255/A18085), through the Cancer Research UK Oxford Centre, grants from the Oxford Wellcome Institutional Strategic Support Fund (204826/Z/16/Z). SPP was partly supported by the National Institute for Health Research

University College London Biomedical Research Centre. None of these organisations was involved in the statistical analysis, interpretation of the results, or writing this article.

Ethical approval

This study utilising QResearch® data has obtained approval from the QResearch Scientific Committee in July 2018. QResearch is a Research Ethics Approved Research Database, confirmed from the East Midlands – Derby Research Ethics Committee (Research ethics reference: 18/EM/0400). A dedicated webpage for this project has been created on the QResearch website https://www.qresearch.org/research/approved-research-programs-and-projects/adepts-accelerated-diagnosis-in-neuroendocrine-and-pancreatic-tumours/. The study protocol and statistical analysis plan are available from this webpage.

Declaration of interests

JHC is an unpaid director of QResearch, a not-for-profit organisation in a partnership between the University of Oxford and EMIS Health, who supply the QResearch database for this work. JHC is a founder and shareholder of ClinRisk Ltd and was its medical director until 31 May 2019. ClinRisk Ltd produces open and closed source software to implement clinical risk algorithms into clinical computer systems. Other authors have no interests to declare for this submitted work.

References

1. Cancer Research UK. Pancreatic cancer statistics [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer. Last accessed date: 23 Feb 2021.
2. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. Lancet. 2016;388(10039):73-85.
3. Clift AK, Kidd M, Bodei L, et al. Neuroendocrine Neoplasms of the Small Bowel and Pancreas. Neuroendocrinology. 2020;110(6):444-76.
4. Kleeff J, Korc M, Apte M, et al. Pancreatic cancer. Nat Rev Dis Primers. 2016;2:16022.
5. Pavel M, O'Toole D, Costa F, et al. ENETS Consensus Guidelines Update for the Management of Distant Metastatic Disease of Intestinal, Pancreatic, Bronchial Neuroendocrine Neoplasms (NEN) and NEN of Unknown Primary Site. Neuroendocrinology. 2016;103(2):172-85.
6. Frilling A, Clift AK. Therapeutic strategies for neuroendocrine liver metastases. Cancer. 2015;121(8):1172-86.
7. Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. World J Oncol. 2019;10(1):10-27.
8. Gedge K. Pancreatic cancer: a symptomless killer. J Perioper Pract. 2017;27(7-8):158-61.
9. Apollos JR, Sami S, Prasanth MN, et al. Pre-diagnostic delays caused by gastrointestinal investigations do not affect outcomes in pancreatic cancer. Ann Med Surg (Lond). 2018;34:66-70.
10. Hippisley-Cox J, Coupland C. Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract. 2012;62(594):e38-45.
11. Collins GS, Altman DG. Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer (Pancreas). Br J Gen Pract. 2013;63(614):e636-42.
12. Richardson DB. An incidence density sampling program for nested case-control analyses. Occup Environ Med. 2004;61(12):e59.
13. Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify men with suspected cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract. 2013;63(606):e1-10.
14. Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract. 2013;63(606):e11-21.
15. Cancer Research UK. Pancreatic cancer symptoms [Available from: https://www.cancerresearchuk.org/about-cancer/pancreatic-cancer/symptoms. Last accessed date: 23 Feb 2021.
16. Pancreatic Cancer UK. Signs and symptoms of pancreatic cancer [Available from: https://www.pancreaticcancer.org.uk/information/signs-and-symptoms-of-pancreatic-cancer/. Last accessed date: 23 Feb 2021.
17. The National Institute for Health and Care Excellence (NICE). Suspected cancer: recognition and referral. NICE guideline [NG12] [Available from: https://www.nice.org.uk/guidance/ng12. Last accessed date: 23 Feb 2021.
18. The National Institute for Health and Care Excellence (NICE). Pancreatic cancer in adults: diagnosis and management [Available from: https://www.nice.org.uk/guidance/ng85. Last accessed date: 23 Feb 2021.
19. Rubin D. Multiple imputation for non-response in surveys. New York, NY: John Wiley; 1987.
20. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7.

21. Stapley S, Peters TJ, Neal RD, et al. The risk of pancreatic cancer in symptomatic patients in primary care: a large case-control study using electronic records. Br J Cancer. 2012;106(12):1940-4.
22. Keane MG, Horsfall L, Rait G, Pereira SP. A case-control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer. BMJ Open. 2014;4(11):e005720.
23. Basuroy R, Bouvier C, Ramage JK, et al. Delays and routes to diagnosis of neuroendocrine tumours. BMC Cancer. 2018;18(1):1122.
24. Maisonneuve P, Lowenfels AB. Risk factors for pancreatic cancer: a summary review of meta-analytical studies. Int J Epidemiol. 2015;44(1):186-98.
25. Bosetti C, Lucenteforte E, Silverman DT, et al. Cigarette smoking and pancreatic cancer: an analysis from the International Pancreatic Cancer Case-Control Consortium (Panc4). Ann Oncol. 2012;23(7):1880-8.
26. Lucenteforte E, La Vecchia C, Silverman D, et al. Alcohol consumption and pancreatic cancer: a pooled analysis in the International Pancreatic Cancer Case-Control Consortium (PanC4). Ann Oncol. 2012;23(2):374-82.
27. Duell EJ, Lucenteforte E, Olson SH, et al. Pancreatitis and pancreatic cancer risk: a pooled analysis in the International Pancreatic Cancer Case-Control Consortium (PanC4). Ann Oncol. 2012;23(11):2964-70.
28. Batabyal P, Vander Hoorn S, Christophi C, Nikfarjam M. Association of diabetes mellitus and pancreatic adenocarcinoma: a meta-analysis of 88 studies. Ann Surg Oncol. 2014;21(7):2453-62.

Box 1 – List of 42 candidate symptoms to be examined in this study

- **Symptoms in the QCancer (Pancreas) model**: abdominal distension (women only), abdominal pain, appetite loss, change in bowel habit, constipation (men only), dysphagia (men only), haematemesis, indigestion, weight loss
- **Symptoms from the literature**: back pain, bloating, diarrhoea, fever/shivering, jaundice, itching/pruritus, nausea, steatorrhoea, vomiting
- **Other potential symptoms to be examined the association with the diagnosis of PDAC/PNEN**: abdominal mass, gastrointestinal bleeding, rectal bleeding, heartburn/gastro-oesophageal reflux (GOR), food regurgitation, flatulence, faecal urgency, dark urine, dry mouth, cracked lips, thirst, loss of taste, night sweats, flushes, rash, paleness, oedema, tiredness, appetite increase, weight gain, bad breath, bruising, sore lips, hypersomnia

Table 1 – Clinicopathological characteristics of study cases with pancreatic cancers

Characteristic		PDAC number (col %)	PNEN number (col %)
	N	23,640	596
Age at diagnosis	**Age bands**		
	25 – 29 years	11 (0.05)	<5
	30 – 39 years	129 (0.55)	25 (4.19)
	40 – 49 years	666 (2.82)	84 (14.09)
	50 – 59 years	2,325 (9.84)	118 (19.80)
	60 – 69 years	5,176 (21.90)	177 (29.70)
	70 – 79 years	7,717 (32.64)	135 (22.65)
	80+ years	7,616 (32.32)	54 (9.06)
Route to diagnosis	Death certificate	85 (0.36)	<5
	Emergency presentation	5,542 (23.44)	78 (13.09)
	GP referral	2,442 (10.33)	147 (24.66)
	Inpatient elective	364 (1.54)	23 (3.86)
	Other outpatient pathway	1,125 (4.92)	112 (18.79)
	Two week wait	1,887 (7.98)	57 (9.56)
	Not recorded	12,195 (51.59)	178 (29.87)
Stage at diagnosis (TNM)	Stage I	274 (1.16)	62 (10.40)
	Stage II	769 (3.25)	53 (8.89)
	Stage III	593 (2.51)	27 (4.53)
	Stage IV	4,016 (16.99)	161 (27.01)
	Not recorded	17,988 (76.09)	293 (49.16)
Grade at diagnosis	Well-differentiated	303 (1.33)	230 (38.59)
	Moderately differentiated	1,683 (7.48)	58 (9.73)
	Poorly differentiated	1,702 (7.52)	61 (10.23)
	Undifferentiated	49 (0.2)	<5
	Not recorded	18,779 (84.19))	246 (41.28)
Type of tumour	Functional NEN	N/A	16 (2.68)
	Non-functional NEN	N/A	271 (45.47)
	Neuroendocrine carcinoma	N/A	274 (45.97)
	Mixed adenocarcinoma/NEN	N/A	8 (1.34)
	Other/NOS	N/A	23 (3.86)
	PDAC	23,640 (100)	N/A
Treatment	Surgery		
	No	17,026 (72.02)	220 (36.91)
	Yes	6,614 (27.98)	376 (63.09)
	Chemotherapy		

	No	19,402 (82.07)	430 (72.15)
	Yes	4,238 (17.93)	166 (27.85)
	Radiotherapy		
	No	23,000 (97.29)	557 (93.46)
	Yes	640 (2.71)	39 (6.54)
	Hormone therapy		
	No	22,547 (99.72)	550 (92.28)
	Yes	66 (0.28)	46 (7.72)
	Other treatment		
	No	19,030 (80.50)	458 (76.85)
	Yes	4,610 (19.50)	138 (23.15)

Note: PDAC – pancreatic ductal adenocarcinoma, PNEN – pancreatic neuroendocrine neoplasm, N/A – not applicable. Rows with counts less than 5 have been suppressed.

Table 2 – Demographics, lifestyle, and comorbidities in cases and controls

Characteristic	Cases with PDAC number (col %)	Cases with PNEN number (col %)	Controls number (col %)
N	23,640	596	230,024
Sex			
Female	11,705 (49.51)	294 (49.33)	114,429 (49.75)
Male	11,935 (50.49)	302 (50.67)	115,595 (50.25)
Age, Mean (SD), years	73.0 (11.5)	62.3 (13.2)	72.0 (11.4)
Median (IQR)	74 (66 – 82)	64 (53 – 72)	73 (65 – 81)
Ethnicity			
White	12,604 (53.32)	367 (61.58)	154,096 (66.69)
Indian	186 (0.79)	9 (1.51)	2,674 (1.16)
Pakistani	100 (0.42)	6 (1.01)	1,150 (0.50)
Bangladeshi	47 (0.20)	<5	692 (0.30)
Other Asian	79 (0.33)	10 (1.68)	1,223 (0.53)
Caribbean	247 (1.04)	10 (1.68)	2,364 (1.03)
Black African	103 (0.44)	<5	1,369 (0.60)
Chinese	36 (0.15)	<5	449 (0.20)
Other	166 (0.70)	12 (2.01)	1,853 (0.81)
Not recorded	10,072 (42.61)	173 (29.03)	64,154 (27.89)
Townsend quintile			
1 (most affluent)	7,054 (29.84)	190 (31.88)	72,988 (31.73)
2	5,873 (24.84)	135 (22.65)	58,901 (25.61)
3	4,758 (20.13)	126 (21.14)	44,581 (19.38)
4	3,442 (14.56)	91 (15.27)	31,513 (13.70)
5 (most deprived)	2,470 (10.45)	54 (9.06)	21,758 (9.46)
Not recorded	43 (0.18)	0 (0.00)	283 (0.12)
Year of index date			
2000-2004	4,358 (18.43)	68 (11.41)	42,402 (18.43)
2005-2009	5,744 (24.30)	107 (17.95)	55,270 (24.03)
2010-2014	6,767 (28.63)	225 (37.75)	66,009 (28.70)
2015-2019	6,771 (28.64)	196 (32.89)	66,343 (28.84)
Smoking status			
Non-smoker	10,841 (45.86)	327 (54.87)	118,626 (51.57)
Ex-smoker	7,117 (30.11)	170 (28.52)	76,842 (33.41)
Light smoker (<10)	3,047 (12.89)	45 (7.55)	16,917 (7.35)
Moderate smoker (10-19)	617 (2.61)	8 (1.34)	2,908 (1.26)
Heavy smoker (20+)	394 (1.67)	7 (1.17)	1,959 (0.85)
Not recorded	1,624 (6.87)	39 (6.54)	12,772 (5.55)
Alcohol consumption			

Non-drinker	14,053 (59.45)	354 (59.40)	139,902 (60.82)
Trivial (<1u/day)	3,224 (13.64)	89 (14.93)	32,028 (13.92)
Light (1-2u/day)	1,480 (6.26)	41 (6.88)	15,194 (6.61)
Moderate (3-6u/day)	1,539 (6.51)	32 (5.37)	14,712 (6.40)
Heavy (7-9u/day)	142 (0.60)	<5	1,173 (0.51)
Very heavy (>9u/day)	67 (0.28)	<5	549 (0.24)
Not recorded	3,135 (13.26)	75 (12.58)	26,466 (11.51)
Comorbidities/previous medical history			
Type 1 diabetes	534 (2.26)	25 (4.19)	1,752 (0.76)
Type 2 diabetes	6,581 (27.84)	187 (31.38)	38,666 (16.81)
Venous Thrombus Embolism	2,631 (11.13)	62 (10.40)	12,565 (5.46)
Deep vein thrombosis	1,632 (6.90)	35 (5.87)	8,246 (3.58)
Pulmonary embolism	1,320 (5.58)	34 (5.70)	5,640 (2.45)
Acute pancreatitis	857 (3.63)	21 (3.52)	2,208 (0.96)
Chronic pancreatitis	383 (1.62)	6 (1.01)	409 (0.18)
Cholangitis	477 (2.02)	8 (1.34)	703 (0.31)
Gallstones	1,602 (6.78)	44 (7.38)	12,378 (5.38)
H. pylori infection	996 (4.21)	37 (6.21)	8,359 (3.63)
Family history of GI cancer	456 (1.93)	15 (2.52)	4,074 (1.77)
Blood cancer	350 (1.48)	21 (3.52)	4,123 (1.79)
Colon cancer	371 (1.57)	10 (1.68)	3,829 (1.66)

Table 3 – Summary of patient and clinical characteristics significantly associated with the diagnosis of PDAC and PNEN

	PDAC	**PNEN**
Demographics	Reduced risks in Indian, Bangladeshi, Other Asian (not including Chinese) compared with white (OR<1)	
Lifestyle	BMI (OR<1) All smoking categories associated with increased risks compared with non-smokers Moderate drinking (3-6 u/day)	
Symptoms (within 3 months)	GI bleeding, Jaundice, Dysphagia (3)	GI bleeding, jaundice (2)
Symptoms (within 1 year)	Diarrhoea, bowel change, vomiting, indigestion, abdominal mass, abdominal pain, weight loss (7) Constipation, steatorrhea, abdominal distension, nausea, flatulence, heartburn, fever, tiredness, appetite loss, itching, back pain, thirst, dark urine (13)	Diarrhoea, bowel change, vomiting, indigestion, abdominal mass, abdominal pain, weight loss (longer than 2 years) (7)
Comorbidities	Type 2 diabetes, Venous Thromboembolism (VTE), Cushing's syndrome, pancreatic cyst acute pancreatitis, cholangitis, family history of GI cancer, Type 1 diabetes	Type 2 diabetes, Venous Thromboembolism (VTE), Cushing's syndrome, pancreatic cyst

Note: PDAC – pancreatic ductal adenocarcinoma; PNEN – pancreatic neuroendocrine neoplasm, those coloured in navy are exclusive to PDAC, and those in black are the same symptoms and comorbidities in both PDAC and PNEN.

Figure 1 – Forest plots for multivariable conditional logistic regression after multiple imputation (alarm symptoms 3 months before the index date, other symptoms 1 year before the index date)

Note: Figure 1-A is for PDAC, and Figure 1-B is for PNEN. Due to a large number of variables in the model, only significant results (adjusted odds ratio and 95% CI) were presented in the figures, considering the figure size for publication space. Full results of the models are available in Supplementary Tables 2-3.

Figure 2 – Forest plots for multivariable conditional logistic regression after multiple imputation (alarm symptoms 6 months before the index date, other symptoms 2 years before the index date)

Note: Figure 2-A is for PDAC, and Figure 2-B is for PNEN. Due to a large number of variables in the model, only significant results (adjusted odds ratio and 95% CI) were plotted in the figures, considering the figure size for publication space. Full results of the models are available in Supplementary Tables 4-5.

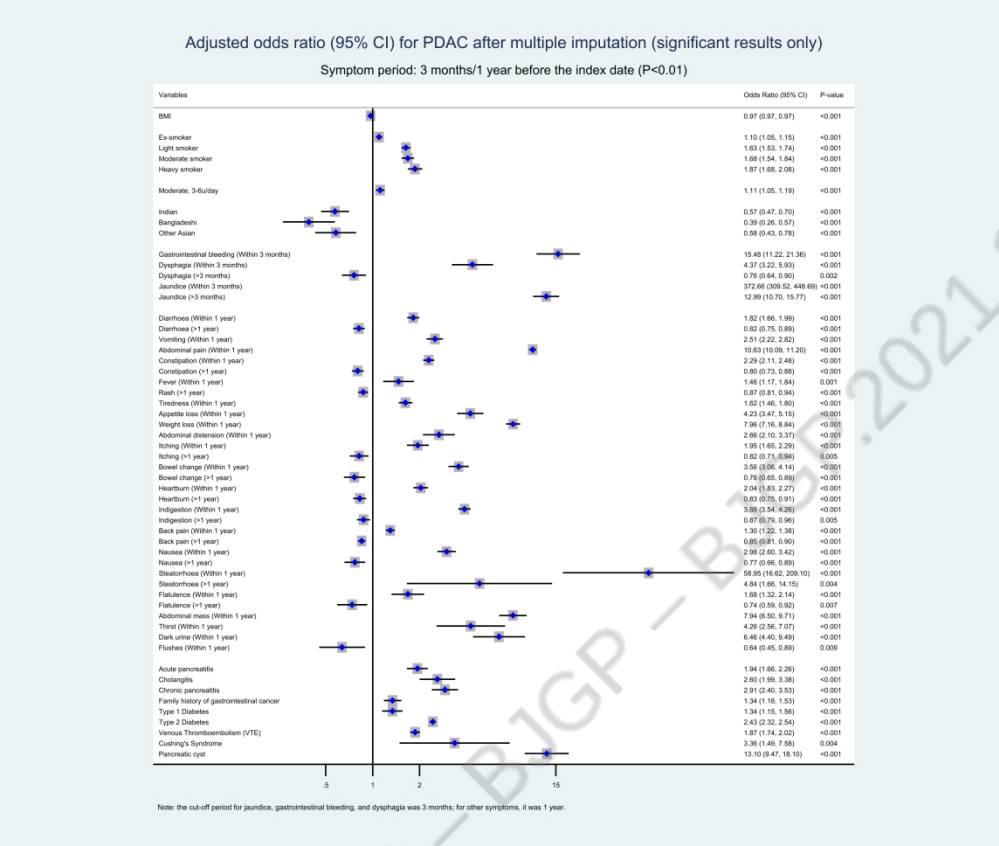

Figure 1-A. Forest plots for multivariable conditional logistic regression after multiple imputation (PDAC, alarm symptoms 3 months before the index date, other symptoms 1 year before the index date)

859x727mm (96 x 96 DPI)

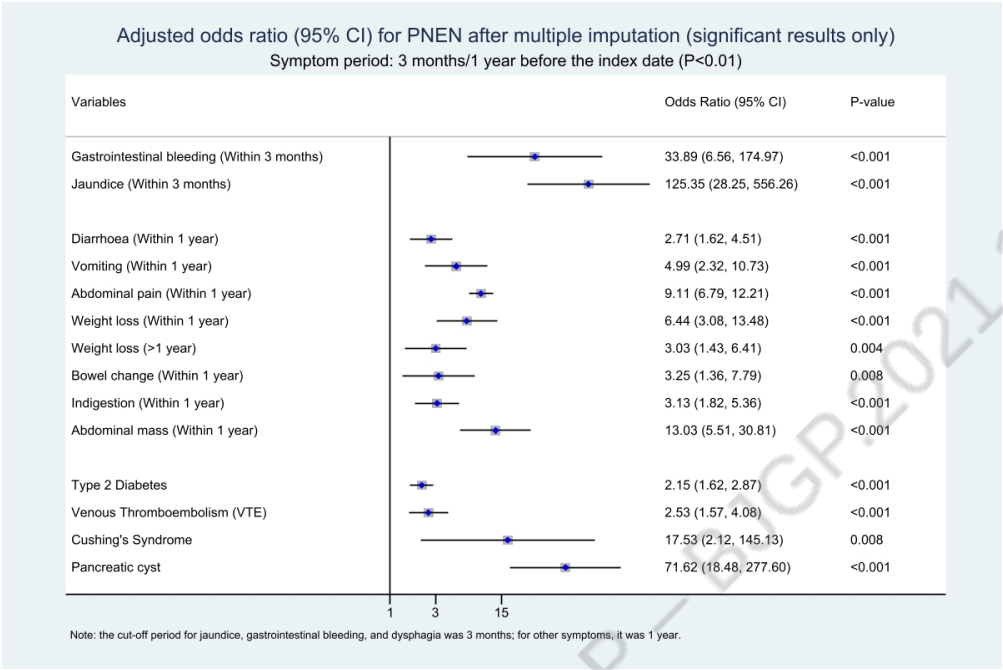

Figure 1-B. Forest plots for multivariable conditional logistic regression after multiple imputation (PNEN, alarm symptoms 3 months before the index date, other symptoms 1 year before the index date)

317x211mm (96 x 96 DPI)

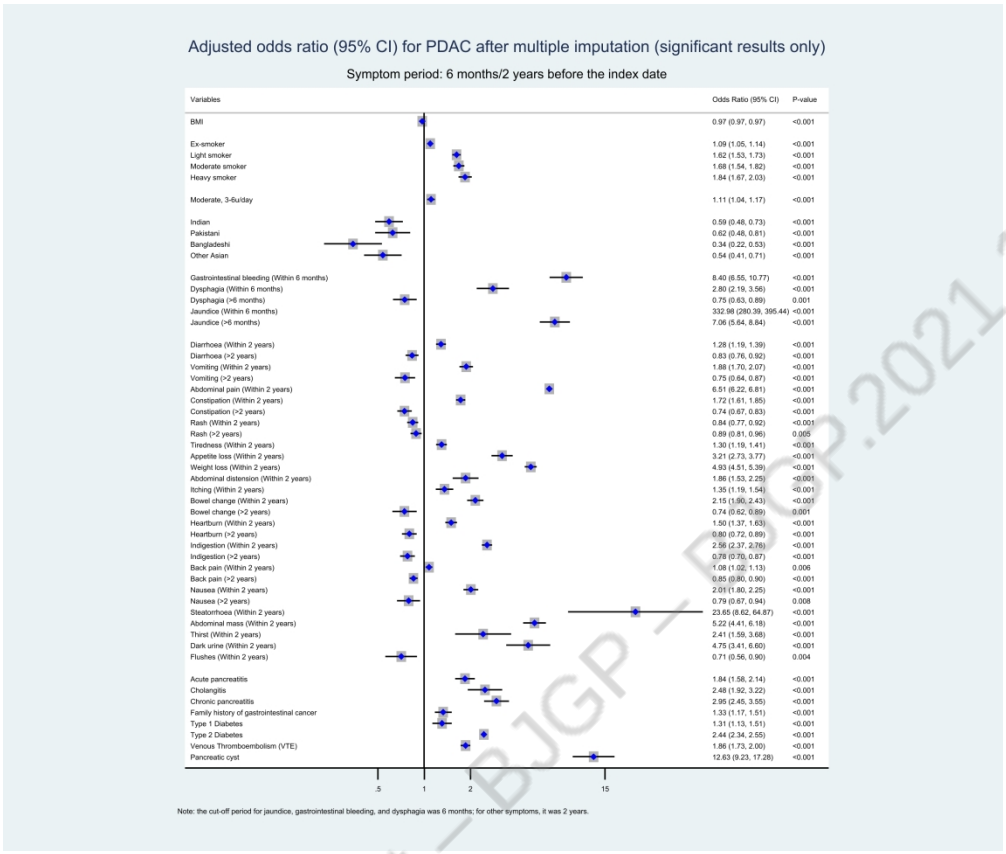

Figure 2-A. Forest plots for multivariable conditional logistic regression after multiple imputation (PDAC, alarm symptoms 6 months before the index date, other symptoms 2 years before the index date)

859x727mm (96 x 96 DPI)

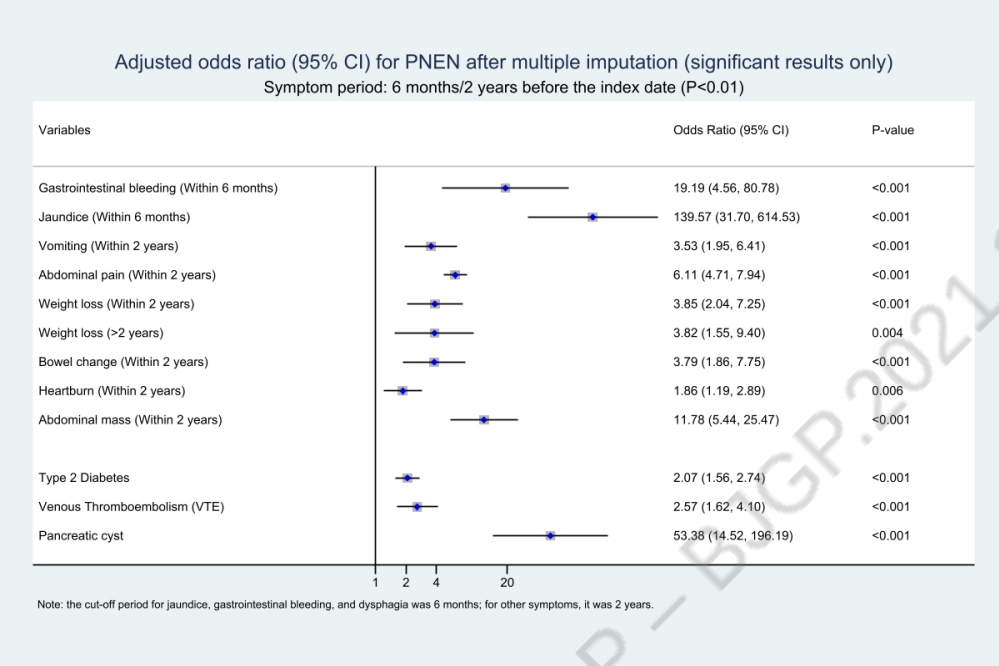

Figure 2-B. Forest plots for multivariable conditional logistic regression after multiple imputation (PNEN, alarm symptoms 6 months before the index date, other symptoms 2 years before the index date)

317x211mm (96 x 96 DPI)